Trending in Pc Measurements via a Bayesian Zero-Inflated Mixed Model

Jonathon Vallejo, Matthew Hejduk, James Stamey

25th ISSFD

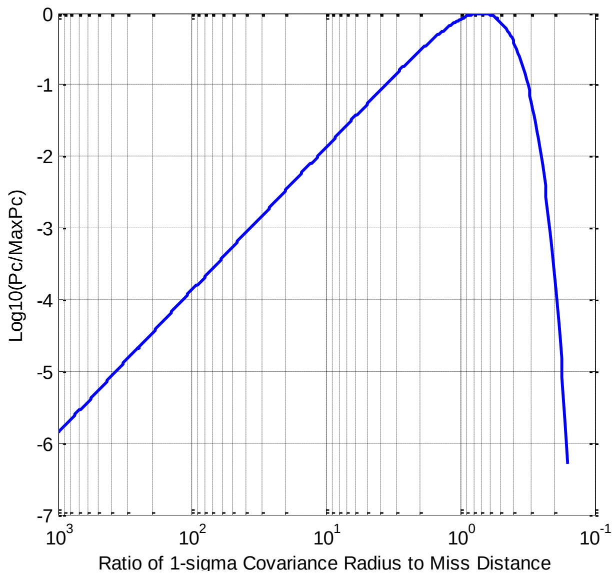
OCT 2015

The Conjunction Assessment Problem

- **Two satellites predicted to come within close proximity of one another**
  - Usually a high-value satellite and a piece of space debris
- **Moving the active satellite is a means of reducing collision risk**
  - But reduces satellite lifetime, perturbs satellite mission, and introduces its own risks
- **So important to get a good statement of the risk of collision in order to determine whether a maneuver is truly necessary**
- **Two aspects of risk statement**
  - Calculation of the Probability of Collision (Pc) based on the most recent set of position/velocity and uncertainty data for both satellites
  - Examination of the changes in the Pc value as the event develops
    - Events in principle should follow a canonical development (Pc vs time to closest approach (TCA))
    - Helpful to be able to guess where the present and future data point fits in the canonical development in order to guide operational response

Conjunction Event Canonical Progression

- **Conjunction usually first discovered 7 days before TCA**
  - Covariances large, so typically Pc below maximum
- **As event tracked and updated, changes to state estimate are relatively small, but covariance shrinks**
  - Because closer to TCA, less uncertainty in projecting positions to TCA
- **Theoretical maximum Pc encountered when 1-sigma covariance size to miss distance ratio is 1/√2**
  - After this, Pc usually decreases rapidly
- **Behavior shown in graph at right**
  - X-axis is covariance size / miss distance (Mahalanobus distance)
  - Y-axis is $\log_{10}(P_c/\max(P_c))$



a.i. solutions

The Pc Progression Problem

- **Information extremely helpful to flight safety operations:**
  - Has the Pc peak been reached?
    - Future Pc values will be only less serious than what has already been observed
    - If observed values not high enough to take action, then event severity reduced
  - Is a presently high Pc likely to fall off by the "maneuver commit point"?
    - Maneuver commit point is time before TCA by which maneuver plans need to be completed and commands sent
    - If reasonable suspicion that Pc will fall off, then events close to remediation threshold less worrisome and need not be worked actively
- **Pc trend models that can answer these questions will contribute significantly to CA operations**

Previous Modeling Effort: Pc Peak Prediction

- **Vallejo, Hejduk, and Stamey 2014 (AAS ASC, Vail, CO)**
- **Modeled Pc time history as inverted parabola**
  - Bayesian framework with informative priors, taken from training dataset
  - Parabolic fit of event data to present given by posterior distribution, calculated through Markov Chain Monte Carlo (MCMC) techniques
- **Used to answer simple question of whether Peak Pc has passed**
  - Correct about 70% of time when tested against entire 2014 dataset
  - Correct about 60% of time against more challenging historical scenarios
- **Not fantastic, but not unpromising results from very simple model**
- **Prompted investigation of more sophisticated model to try to improve performance**
  - Try to predict the actual Pc value at a future point
    - Could be used to decide to cease active working of certain events

Modeling Choice of Variables

- **Dependent variable is log10 value of Pc**
  - Significant changes in Pc on the "order of magnitude" level, thus $\log_{10}$Pc
  - Need to address problem of very small and 0 values for Pc
  - For purposes of operations Pc values "essentially 0" when less than 1E-10
    - Small values of Pc can thus be "floored" at 1E-10
    - Long trains of leading or trailing 1E-10 values can also be eliminated from dataset; really just a function of when updates happen to occur.
- **Independent variable is time before TCA (usually in fractional days)**
  - Canonical behavior curve uses independent variable as ratio of covariance size to miss distance
  - Problematic independent variable for model
    - Not monotonic with time (but it does correlate at least moderately to time)
    - Need temporal independent variable in order to map to operational timelines
  - Thus, use time before TCA as independent variable for model

Distribution Choice for $\log_{10} P_c$ Modeling

- **Usual choice for bounded random variables is Beta distribution**
  - Because $\log_{10} P_c$ values floored at -10, have bounded $\log_{10} P_c$ values between -10 and 0
- **When scaled to (0,1) interval, -10 values will become zeroes**
- **Unmodified beta distribution cannot actually accommodate zero values**
  - Extension to allow this creates a “zero-inflated” beta distribution:

$$f(\mathrm{y} \mid \mu, \phi, p) = (1-p)\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} + pI_{[0]}(\mathrm{y})$$

  - Core portion of distribution is first term
  - Indicator function is second term (sets value equal to zero)
  - P is the probability that the distribution will yield a zero

What about μ and p?

$$f(y \mid \mu, \phi, p) = (1-p)\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} + pI_{[0]}(y)$$

- **These parameters, which express the mean and the zero-inflation probability, can be single parameters or linear functions**
- **In a mixed-model framework, they can also include random elements**
  - Better way to specify overall trend yet random effects for each event
- **Trial runs with training dataset indicated that a second-order linear model with a random intercept (constant) produced best results (minimum deviance; see paper)**
- **Parameterized functions for μ and p are thus as follows:**

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + b_i \qquad \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha_0 + \alpha_1 t_{ij} + \alpha_2 t_{ij}^2 + a_i$$

Bayesian Inference Basics

- **We solve not for single parameter values but the posterior density of the parameters given the data**
  - Suppose we have a model with parameter β
  - $p(\beta|y) \propto p(y|\beta) * p(\beta)$
  - Thus, we specify a prior distribution for the parameter β p(β), update it with the data that we have seen p(y|β), and get an updated probability distribution of beta given the data p(β|y).
- **Prior (here historical) information included through the use of informative prior distributions**
- **Posterior density is thus combination of trends derived from prior information and event-specific information up to the point from which predictions are to be made, as in the following example scenario**
  - Informative priors come from last years' conjunction information database
  - Current event information is actual Pc values from 7 through 4 days to TCA
  - Posterior distribution prediction is of the Pc value at 2 days to TCA

Model Parameter Assigned Distributions

Let Y_{ij} be the predicted log(P_c) at the j^{th} time for the i^{th} event, scaled to be between 0 and 1.

$$Y_{ij} \sim f(y_{ij} \mid \mu_{ij}, \phi_{ij}, p)$$

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + b_i \qquad \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha_0 + \alpha_1 t_{ij} + \alpha_2 t_{ij}^2 + a_i$$

$$b_i \sim N(0, \tau_b) \qquad a_i \sim N(0, \tau_a)$$

$$\tau_b \sim Gamma(0.001, 0.001) \qquad \tau_a \sim Gamma(0.001, 0.001)$$

$$\beta_k, \alpha_k \sim Normal(0,1) \quad \text{for } i = 1,2,3$$

Mixed Model Comments

$$f(\mathrm{y} \mid \mu, \phi, p) = (1-p)\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} + pI_{[0]}(\mathrm{y})$$

- ***μ*** **and** ***p*** **both have a linear and a random portion**
  - Linear portion (here quadratic) expresses trends across entire dataset
  - Random portion expresses observed behavior of current event in progress
- ***μ*** **specifics**
  - The model for μ is a model for the average of the log(Pc) values that fall *between* -10 and 0.
    - A positive random intercept indicates that one is more likely to see a higher than usual log(Pc) value in the subsequent days.
- ***p*** **specifics**
  - The model for *p* is a model for the probability of observing a log(Pc) equal to -10 (i.e. a Pc equal to 0).
    - A positive random intercept indicates that one is more likely to observe a log(Pc) of -10 than usual.

φ and t Comments

$$f(y \mid \mu, \phi, p) = (1-p)\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} + pI_{[0]}(y)$$

- **φ controls the variance in the data**
  - This parameter found to perform best as a constant, thus no linear model associated with it
  - Model naturally accommodates the changing variability, so no extra model needed for φ
- **Model uses only time (*t*) as a covariate**
  - Thus only three pieces of information informing the model:
    1. How high/low the log(Pc) values are relatively to the average
    2. How many log(Pc) values equal to -10 have been observed relative to the average
    3. How unusual these observations are at the particular moment in time they were taken (relative to TCA)

Model Training Procedures

- **Employed 2013 NASA conjunction data for 500-750km orbits**
  - “Training” dataset
- **Used quite uninformative priors**
  - Large variances so that both common and extreme data can be represented
  - Allows full dataset to influence results
- **Posterior distributions should thus incorporate statistical properties of the actual data**
  - How training dataset behaves and develop over time a template of what to expect from future data
- **These posterior distributions become the prior distributions for the model when it is run against validation data**

Beta Model Overall Commentary (1 of 2)

- **Advantages of model:**
  - Random intercept used to quantify how much $\log(P_c)$ values from any event deviate from the overall mean
    - For instance, if an event had a really high value of a_i (the random intercept in the linear model for *p*), we would interpret this as a higher than average chance of getting a zero during this event
  - The model more closely follows the overall shape of the data. The model includes the *p* parameter, which can be directly interpreted as the probability of getting a P_c of 0 (or a $\log(P_c)$ of -10)
  - The beta model accommodates non-constant variance
    - If the $\log(P_c)$ values are closer to 0 (*i.e.,* the P_c values are closer to 1), then the variance is relatively small
    - Likewise, if the $\log(P_c)$ values are closer to -10 (*i.e.,* the P_c values are closer to 0), then the variance is relatively large
  - The model can easily be made more conservative. If one wants to upwardly bias the predictions, simply choose the 75^{th} or 97.5^{th} quantiles of the random intercept instead of its mean

Beta Model Overall Commentary (2 of 2)

- **Disadvantages of model**
  - As a result of the shape of the overall mean, the maximum predicted $\log(P_c)$ value is always at 7 days from TCA
  - This drives how model can be deployed most usefully
    - Not helpful method for peak prediction
    - But well suited to predict drop-offs in Pc value
    - As such, should be able to identify cases that are likely to become non-threatening
    - Will not model truly anomalous Pc progressions

“Foil” for Beta Model Evaluation: Naïve Look-up Model based on Quantiles

- **Simplest method for predicting future of time-series events for which there are historical data**
  - If unfolding event is in a certain historical quantile at a given time t, then it can be expected to remain in the same quantile at time t_n
  - Training dataset can thus be used to estimate Pc at t_n
- **Represents, to first order, how many analysts intuitively make decisions**
  - Event of a certain severity at present time is likely to be of an expeted different severity at a given future time
- **Has an attractive simplicity**
- **Also has certain drawbacks**
  - No real theory standing behind it—why would it be true that historical Pc histories would be stratified in this way?
  - No inherent prediction intervals because no distributional assumptions made
    - These must come from bootstrapping techniques
- **If Beta model cannot outperform this, then relevance questioned**

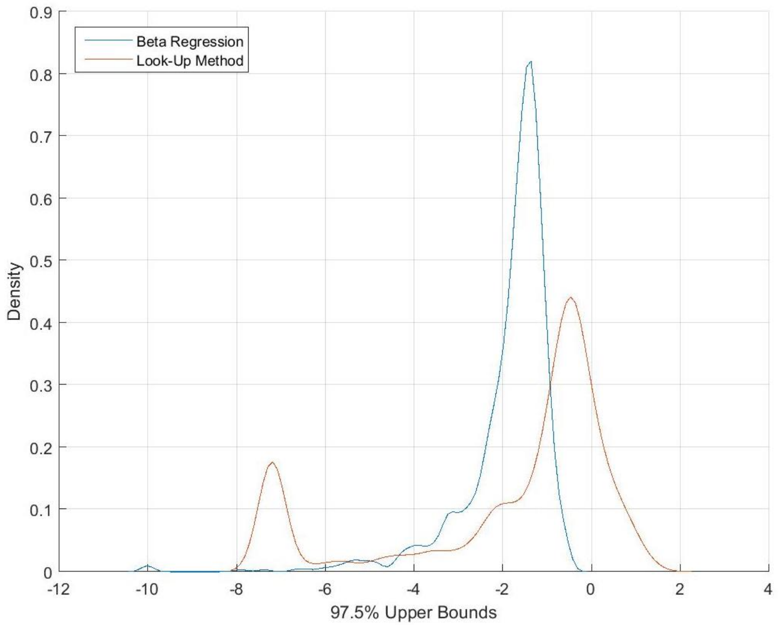
a.i. solutions

Model Validation Dataset

- **2014 NASA CA conjunction message database**
- **LEO2 orbits (500-750 km, near-circular primaries)**
- **Early coverage assessment for beta model revealed difficulties**
  - 86% actual coverage for 97.5% prediction
  - Suggested data stratification as a remedy
- **Data divided into three strata, based on operational severity**
  - Events with Pc > 1E-04 at three days before TCA ("red" events)
  - Events with Pc between 1E-07 and 1E-04 at three days before TCA ("yellow")
  - Events with Pc < 1E-07 at three days before TCA ("green")
- **Coverage improves substantially when different strata processed separately**
  - For example, red dataset produces coverage levels of 97.6% and 97.4% for beta and look-up models, respectively—both excellent
  - Stratified datasets used for remainder of validation activities

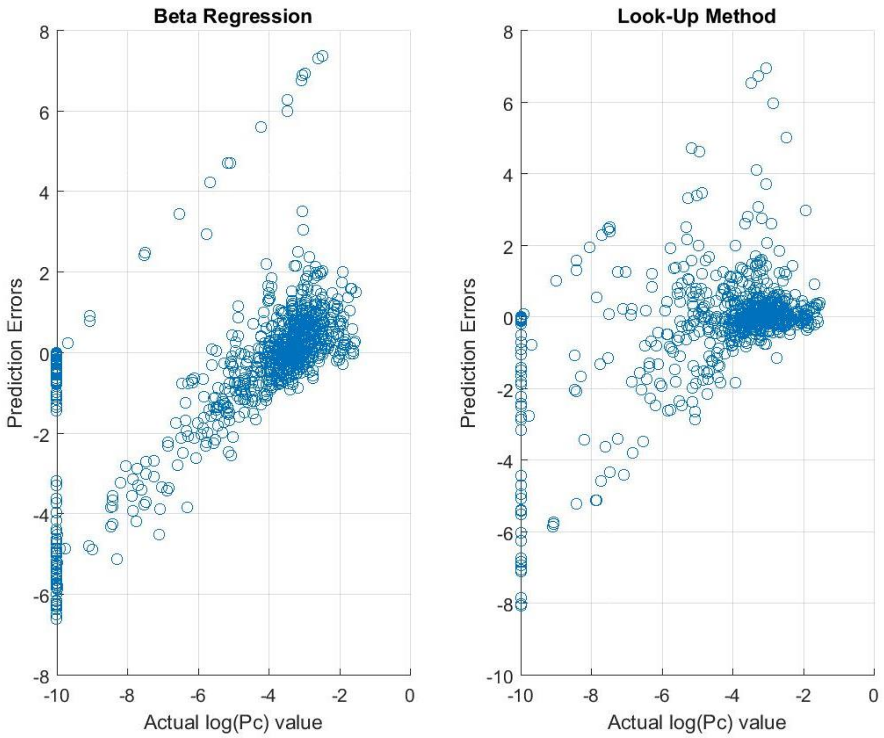
Prediction Probability Densities (Red Dataset)

- **Shown in graph at right**
- **Look-up model has desired bimodality**
  - Peak for prediction of high Pc and second peak for prediction of drop-off to zero
  - However, bootstrap technique produces physically impossible results (Pc>1)
- **Beta model remains within desired bounds**
  - However, not well poised to predict drop-offs to zero, as very little probability density in this region



Prediction Residual Errors (Red Dataset)

- **Both models have positive bias**
  - Nature of data: cannot predict a value below -10, so overpredict
- **Quantile model more symmetric and bounded**
- **Beta model has systematic effects and weaker performance**
  - Somewhat disappointing result
- **Both models struggle with predicting the drop-offs to zero**
  - Although quantile model performs more strongly

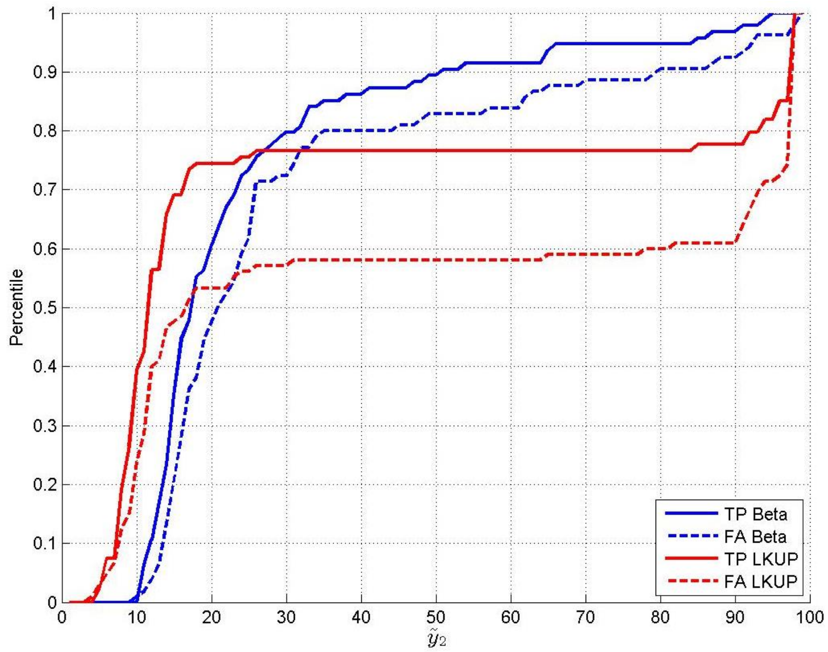


ROC Evaluation of Yellow Dataset

- Receiver Operating Characteristic (ROC) curves useful for evaluating decision support classification algorithms
- True Positive event: here defined as the correct identification of an event where the Pc will remain high
- False Alarm event: here defined as the incorrect flagging of a drop-off event as a "remaining high" event
- Missed Detection event: 1 – True Positive level
- Plots give ROC CDF as a function of upper percentile of both beta and look-up models

Yellow Dataset ROC Results

- **At lower percentile levels, look-up model performance superior**
  - More true positives with lower false alarm rate
- **At upper percentiles, situation reversed**
- **Operational utility requires a very high true positive rate**
  - Minimizes Type II errors
- **Thus, look-up model not useful here**
- **Beta model could be useful, but false alarm rate (Type I errors) very high**



Results and Future Work

- **Peak identification capability (from previous effort) provides limited but palpable benefit**
- **Quantile approach provides small utility for red dataset**
- **Beta approach provides small utility for yellow dataset**
- **Emerging conclusion**
  - Trending approaches can provided limited additional operational information
  - Not likely to be a breakthrough or transformative technology for conjunction assessment
  - Will need to determine proper role of such tools within operational decision support framework
- **Future work**
  - One more trending method to explore—functional/longitudinal data analysis
  - Omnibus evaluation of all four methods investigated